Biodiversity of pollen in indoor air samples as revealed by DNA metabarcoding

Korpelainen, Helena

2017-10

Korpelainen , H & Pietiläinen , M 2017 , ' Biodiversity of pollen in indoor air samples as revealed by DNA metabarcoding ' , Nordic Journal of Botany , vol. 35 , no. 5 , pp. 602-608 . https://doi.org/10.1111/nj

http://hdl.handle.net/10138/307916
https://doi.org/10.1111/njb.01623

acceptedVersion

Revision

Biodiversity of pollen in indoor air samples as revealed by DNA metabarcoding

Helena Korpelainen* and Maria Pietiläinen

Department of Agricultural Sciences, Viikki Plant Science Centre, P.O. Box 27, FI-00014 University of Helsinki, Finland

*Corresponding author

E-mail: helena.korpelainen@helsinki.fi

Phone: +358 29 4158383

We conducted DNA metabarcoding (based on the nuclear ITS2 region) to characterize indoor pollen samples (possibly accompanied by other plant fragments) and to discover whether there are seasonal changes in their taxonomic diversity. It was shown that DNA metabarcoding has potential to allow a good discovery of taxonomic diversity. The numbers of spermatophyte families and genera varied greatly among sampling sites (pooled results per building) and times, between 9-40 and 10-66, respectively. Comparable Shannon's diversity indices equaled 0.33-2.76 and 0.94-3.16. The total number of spermatophyte genera found during the study was 187, of which 43.9, 39.6, 7.5 and 9.1% represented wild, garden/crop and indoor house plants, and non-domestic fruit or other plant material, respectively. Comparable proportions of individual sequences equaled 77.4, 18.8, 2.7 and 1.1%, respectively. When comparing plant diversities and taxonomic composition among buildings or between seasons, no obvious pattern was detected, except for the second summer, when pollen coming from outdoors was highly dominant and the proportions of likely allergens, birch, grass, alder and mugwort pollen, were very high. The average pairwise values of $Sørensen_{Chao}$ indices that were used to compare similarities for taxon composition between samples among the samples from the two university buildings, two nurseries and farmhouse equaled 0.514, 0.109, 0.564, 0.865 and 0.867, respectively, while the mean similarity index for all samples was 0.524. Cleaning frequency may strongly contribute to the observed diversity. The discovery of considerable diversities, including pollen coming from outside, in both winter and summer shows that substantial amounts of pollen produced in summer enter buildings and stay there throughout the year.

Keywords: DNA metabarcoding, indoor air quality, next generation sequencing, plant diversity, pollen, seasonal variation, taxonomic composition

Particles of indoor environments, such as dust mites, mold, bacteria, viruses, pet dander and pollen, are major environmental concerns for human wellbeing, as they can cause severe health problems (Chapman et al. 2007). Some of them, like viruses and bacteria, cause infections, while others cause allergies in susceptible persons, possibly very serious ones. There are estimates that nowadays allergic diseases caused by plant, animal and fungal allergens affect even more than 30% of the population in industrialized countries (Crameri et al. 2014).

Pollen can enter buildings through open windows and doors, and people track pollen indoors on their shoes, clothes and hair. Pollen counts are higher in the spring and summer, although it can remain indoors through other seasons as well. Pollen of most plant species has some level of allergenicity but some types are particularly notorious for inducing symptoms of hay fever. In Finland, in the area of this study, the most problematic types of pollen are those of grasses (Poaceae), birch (*Betula* sp.), alder (*Alnus* sp.) and mugwort (*Artemisia vulgaris*) (Jantunen et al. 2012). Current pollen monitoring methods are microscope-based and labor-intensive. Although pollen of each taxon has its own unique set of characteristics, it is very time-consuming and sometimes impossible to comprehensively determine the taxonomic composition of these tiny particles in air samples without molecular tools (Khansari et al. 2012, Galimberti et al. 2014).

Recent advances in DNA sequencing provide effective tools for species identification and biomonitoring using DNA present in the environment. DNA barcoding through high-throughput sequencing (next generation sequencing) allows the characterization of the species composition of bulk samples, including both intact and degraded DNA extracted from environmental samples (e.g. Taberlet et al. 2012, Bohmann et al. 2014), for example investigations on honey bee pollen foraging and honey composition (Galimberti et al. 2014, Bruni et al. 2015, Cornman et al. 2015, Hawkins et al. 2015, Keller et al. 2015), vegetation analyses in lake sediments (Parducci et al. 2013) and pollen monitoring in air (Kraaijeveld et al. 2015). Such DNA metabarcoding uses

universal PCR primers to mass-amplify a taxonomically informative gene from bulk samples. Recently, we analysed fungal diversity in indoor air by DNA metabarcoding (Korpelainen et al. 2016, Korpelainen and Pietiläinen 2017) and now extend the analysis to plant particles.

In the present study, our aim was to increase precision in analyses on pollen and plant fragments and to provide useful data and tools for investigations on the quality of indoor spaces. Our additional goal was to discover, whether there are seasonal changes in the biodiversity of plant materials in indoor spaces. We chose to use the nuclear ITS2 region, because prior investigations support its universal nature (good PCR amplifiability across taxa) and suitability for differentiating plant taxa in pollen samples at the genus and, in some cases, at the species level (e.g. Richardson et al. 2015a, Sickel et al. 2015), although there is also indication that plastid markers, such as matK and and trnH-psbA, may be more effective when characterizing the diversity of pollen samples (Richardson et al. 2015b).

Material and methods

Samples were collected from five buildings, including two university buildings, two nursery schools and an old farmhouse. The farmhouse is located in the county of Porvoo, about 40 km to the east from Helsinki, while other buildings are located in Helsinki (about 60°14’ N, 25°01’ E). The distance between the two nursery schools is about 6 km, and the distance between the two university buildings is 0.2 km, while the distances between each nursery school and both university buildings are about 12 and 7 km, respectively. The large university buildings and the single-floor nursery school buildings are surrounded by lawns, bushes and primarily broad-leaved trees. The farmhouse is surrounded by a lawn, kitchen garden and fields of cereal crops, and there are many types of trees nearby. Sampling was conducted four times: January 2013, July 2013, January 2014, and July 2014. All buildings were not sampled on every occasion (Table 1). Both nursery schools were renovated during the study due to minor water damage and

observed mould growth, and we sampled them both before and after renovation. Indoor sampling was conducted using a collector with a disposable filter (DUSTREAM™ Collector, Indoor Biotechnologies Inc., Charlottesville, VA, USA; mesh size 40 µm) attached to the tube of a vacuum cleaner with the suction power of 32 L/s. Both a horizontal (tables or shelves) and vertical (walls) sample were collected by vacuuming an area of about 2 m^2/sample (i.e., two 2 m^2 samples per room) from two rooms in each of five buildings (two office rooms in each university building, two playrooms in each nursery school, and two bedrooms in the old farmhouse).

After vacuuming, the filter containing the dust was removed from the collector and placed in a plastic bag until processing, involving cutting the filter, rinsing the filter with water and emptying the content to a petri dish, where large non-biological particles were removed. Thereafter, the samples were dipped in liquid nitrogen and ground in a ball mill, and DNA was extracted using the CTAB method (Doyle and Doyle 1987). The final volume was 100 µl.

For DNA metabarcoding, genomic ITS2 sequences were amplified and sequenced using two approaches. All sequencing was conducted at the DNA Sequencing and Genomics Laboratory, Institute of Biotechnology, University of Helsinki. The sequencing for the samples from January and July 2013 were conducted using 454 FLX pyrosequencing (Roche Applied Science, Penzberg, Germany), as described in Korpelainen et al. (2016), while sequencing for the samples from January and July 2014 were performed using Illumina MiSeq sequencing (San Diego, CA, USA), for which ITS2 sequences were first amplified using the following primer system (forward ITS4 mix + reverse ITS3 mix) (see Korpelainen and Pietiläinen 2017):

Forward ITS4 mix including three primers:

ITS4_F1 5'-ATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTCCTCCGCTTATTGATATGC-3'

ITS4_F2 5'-ATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT(c/g)TCCTCCGCTTATTGATATGC-3

ITS4_F3 5'-ATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTagt(a/g)(a/g)TCCTCCGCTTATTGATATGC-3'

Reverse ITS3 mix including three primers:

ITS3_R1 5'-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGCATCGATGAAGAACGCAGC-3'

ITS3_R2 5'-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT(c/t)GCATCGATGAAGAACGCAGC-3'

ITS3_R3 5'-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTa(a/t)GCATCGATGAAGAACGCAGC-3'

All 20-µl PCR reactions contained 2 µl of template DNA. After sequencing, primer sequences were removed from the raw reads, and quality control, as described by Brown *et al.* (2013), followed. During this process, low-quality reads (below average PHRED score of 25) and short sequences (< 100 bp) were removed. Then, all other sequence data were subjected to similarity search against GenBank (www.ncbi.nlm.nih.gov/genbank), and assignment of taxonomic identities using TAXAassign (https://github.com/umerijaz/ taxaassign) was conducted with 95 and 97% thresholds for genus and species levels, respectively. The 97% threshold is, by convention, used as a divergence threshold for operational taxonomic units (OTUs) that serve as a proxy for species (Brown ***et al.*** 2015). The sequence data were submitted to the EMBL (European Molecular Biology Laboratory) database under accession number PRJEB8345. Based on the numbers of sequences representing each taxon (i.e., taxon distribution), Shannon's diversity indices (Shannon 1948) were calculated at family and genus levels for each sample. In addition, EstimateS 9.1.0. (http://purl.oclc.org/estimates) was used to calculate similarities for taxon composition between pooled longitudinal samples (each including all four samples from a building at the same time point). The used estimator was the $Sørensen_{Chao}$ abundance-based similarity index (corrected for unseen shared species), which can also handle different sample sizes (Chao et al. 2005).

Results

As a result of DNA metabarcoding, good-quality sequences were obtained. Small-scale pyrosequencing was conducted for the two first sets of samples (winter 2013 and summer 2013), and the total number of good sequences averaged 26 276 and 19 868 sequences/building. However, the majority of the sequences represented fungi, and the average numbers of spermatophyte sequences equalled 613 (range 109-1 493) and 537 (range 249-726) sequences/building. For the last two sets of Illumina-sequenced samples (winter 2014 and summer 2014), the total numbers of good sequences were on average 855 576 and 2 235 022 sequences/building, respectively, while the mean numbers of spermatophyte sequences among them equalled 664 (range 399-1041) and 5934 (range 482-10953), respectively.

It is notable that there were no major changes in spermatophyte sequence numbers between winter samples analysed with different sequencing approaches, while Illumina-based analyses in summer 2014 revealed high numbers of sequences, 6360 and 10954 sequences in Nursery 1 and 2, respectively, which equal to 10-100 -fold increases compared to sequence numbers in 2013 when 454 FLX pyrosequencing had been used. However, the third building, University 1, analysed in summer 2014, showed an even slightly lower number of sequences (change from 726 to 482 sequences between summers). The used method is effective until the genus-level identification of spermatophytes (99.7%), but ITS2 alone is not satisfactory for the species-level identification (only 18.0% of the samples). Here, we present diversity and taxonomic information based on genus level data unless specified differently.

The numbers of spermatophyte families and genera per sample varied greatly among sampling sites (pooled results per building) and times, between 9-40 and 10-66, respectively (Table 1). Comparable ranges of Shannon's diversity indices were 0.33-2.76 and 0.94-3.16, respectively. The total number of spermatophyte genera found during the study was 187, of which 43.9, 39.6, 7.5 and 9.1% represented wild, garden/crop plants, indoor house plants, and

non-domestic fruit or other plant material, respectively. Comparable proportions of individual sequences equaled 77.4, 18.8, 2.7 and 1.1%, respectively. Proportions of these four groups of plants, based on sequences numbers, showed great variation among buildings and seasons (Fig. 1). In the last samples from July 2014, a great majority, almost 100% of sequences, represented pollen coming from outdoors (wild and garden/crop plants); also the proportions of sequences representing likely allergens, birch, grass, alder and mugwort pollen, were then very high, 58.5% (24.1% in the whole dataset). Both the winter and summer samples of the farmhouse possessed very high proportions of outdoors pollen material (94.8 and 99.3% of sequences, respectively).

The change in the sequencing method and differences in numbers of sequences between the first two and last two sampling times (particularly summer 2014) did not show in diversity indices (genus level, determined for individual samples) that equalled 1.82±0.70 and 1.59±0.90 ($t=0.766$, $p=0.451$, $df=26$; t test), respectively. However, concerning the numbers of detected taxa, there was a significant increase, the genus numbers equalling 21.4±13.0 and 34.6±15.6 ($t=2.452$, $p=0.021$, $df=26$), respectively. The numbers of taxa were closely similar among individual horizontal and vertical samples, on average 26.2±16.7 and 27.5±14.5 $t=0.211$, $p=0.834$, $df=26$), respectively, as also the diversity indices, on average 1.65±0.91 and 1.81±0.7 ($t=0.536$, $p=0.597$, $df=26$), respectively. Comparable values for individual winter and summer samples showed that there was a tendency to a higher diversity in summer, the values equalling 21.3±14.1 and 31.6±15.2 ($t=1.8657$, $p=0.073$, $df=26$), respectively, and 1.63±0.58 and 1.81±0.93 ($t=0.619$, $p=0.541$, $df=26$), respectively.

Table 2 lists the five most frequent plant taxa detected in each building at each sampling time. The results show that there was a great turnover in the proportions of different taxa. *Brassica* sp. was found among top-five taxa in 10 out of 14 building samples (detected in all building samples), *Betula* sp. in 9 out of 14 building samples (detected in 12 out of 14 building samples), and the common house plant *Saintpaulia* sp. and the common garden ornamental

194 *Syringa* sp. both were among top-five taxa in 4 out of 14 buildings (Table 2). In several samples,
195 one specific taxon was highly dominating, such as *Aegopodium* sp. in University 1 in winter
196 2013 (56.5%), *Brassica* sp. in University 2 in winter 2013 (55.0%), *Fallopia* sp. in University 2
197 in summer 2013 (64.2%), *Ficus* sp. in Nursery 1 in summer 2013 (52.4%), *Syringa* sp. in
198 Farmhouse in winter 2014 (83.3%), and *Betula* sp. in Nursery 1 and Nursery 2 in summer 2014
199 (47.9 and 97.9%, respectively).

200 Sørensen$_{Chao}$ indices that were calculated to compare similarities for taxon composition
201 between samples did not show any clear pattern. The average pairwise values for temporal
202 pooled samples among University 1, University 2, Nursery 1, Nursery 2 and farmhouse samples
203 equaled 0.514, 0.109, 0.564, 0.865 and 0.867, respectively, while the mean similarity index for
204 all samples was 0.524. The similarity indices of Nursery 1 and Nursery 2 samples for before and
205 after renovation samples equaled 0.333 and 0.865, respectively. For comparison, Shannon's
206 genus-level diversity indices of Nursery 1 equaled 1.80 and 1.59 before and after renovation, and
207 those of Nursery 2 equaled 2.53 and 1.01, respectively.
208
209 **Discussion**
210 DNA metabarcoding was conducted successfully for pollen samples (possibly accompanied by
211 other plant particles) collected from indoor spaces in five buildings, including two university
212 buildings, two nursery schools and an old farmhouse. The change of sequencing method from
213 454 FLX pyrosequencing (winter 2013 and summer 2013 samples) to Illumina MiSeq
214 sequencing (winter 2014 and summer 2014 samples) and resulting increases in sequence
215 numbers (especially for summer 2014 samples) are suggested to increase the detection of
216 infrequent taxa, with the mean number of genera per building increasing from 22 to 35.
217 However, temporal variation, partly due to human effects, such as cleaning frequency, may
218 contribute to changes in numbers of sequences and detected taxa. Considering sequencing

219 platform qualities, Kozich et al. (2013) have demonstrated that Illumina MiSeq platform can
220 provide data that are at least as good as that generated by the 454 platform while providing
221 higher sequencing coverage at a lower cost. Nelson et al. (2014) have also discussed how well
222 Illumina sequencing could serve as a direct replacement for 454 pyrosequencing. They showed
223 in diversity analyses on microbial communities based on the 16S region that moving to Illumina-
224 based sequencing platforms provides deeper insights into the breadth of diversity, but they
225 pointed out that care must be taken to ensure that sequencing and processing artefacts do not
226 obscure the results. Recently, Castelino et al. (2017) have also shown that Illumina provides
227 comparable data to 454 pyrosequencing, with a similar capture of diversity but with a much
228 improved throughput and cost effectiveness.

229 Plant diversities in samples collected from different buildings (university offices, nursery
230 schools and a farmhouse) and during different seasons (summer vs. winter) showed considerable
231 variation and turnover but no definite pattern, although there was a slight tendency to a higher
232 diversity in summer. Also, $Sørensen_{Chao}$ indices that were calculated to compare similarities for
233 taxon composition between samples did not show any clear pattern. The mean similarity index
234 for all samples was 0.524, while the similarity indices of Nursery 1 and Nursery 2 samples for
235 before and after renovation samples equalled 0.333 and 0.865, respectively. For comparison,
236 Shannon's genus-level diversity indices of these nursery samples were lower after renovation,
237 which may relate to extensive cleaning of buildings after renovation.

238 The proportions of wild, garden/crop plants, indoor house plants, and non-domestic fruit
239 or other plant material, based on sequences numbers in our analyses, showed great variation
240 among buildings and seasons. In several samples, one specific taxon was highly dominating. In
241 the last sample set from July 2014 almost 100% of sequences represented plants coming from
242 outdoors (wild and garden/crop plants), and the proportions of sequences representing likely
243 allergens, birch, grass, alder and mugwort pollen, were then very high (58.5%). Both the winter

and summer samples of the farmhouse possessed very high proportions of outdoors plant material (94.8 and 99.3%), which may relate to the rural setting and perhaps even more to the lack of air conditioning and frequent opening of windows for cooling and air renewal, thus facilitating the penetration of pollen from outside.

Considerable variation in pollen (plant) composition found to occur even within the same building emphasizes the importance of multiple sampling. A considerable presence of certain allergy-inducing plants, such as birch (*Betula* sp.) and grasses (Poaceae), may be good indicators of indoor air quality (as far as plant particles are concerned) and may indicate the need of improved cleaning or air purification. Besides seasonality, cleaning frequency and coverage in different buildings may strongly contribute to the observed diversity and pattern of plant particles, although all studied buildings are cleaned regularly. The discovery of considerable diversities, including also pollen coming from outside, in both winter and summer shows that substantial amounts of pollen produced in summer enter buildings and stay there throughout the year, as previously observed also by Pichot et al. (2015).

DNA metabarcoding is considered as an effective tool for biodiversity investigations, and its effectiveness is already well proven for a range of organisms and environments, and even for poor-quality and low-quantity DNA (e.g. Taberlet et al. 2012, Bohmann et al. 2014, Korpelainen et al. 2016, Korpelainen and Pietiläinen 2017). The approved standard barcode for the land plants is a two-locus DNA barcode, including a portion of coding chloroplast genes rbcL and matK (CBOL Plant working group 2009). However, for the purpose of our analysis, a single-region approach was more practical. While rbcL does not provide satisfactory species discrimination power (CBOL Plant working group 2009), matK is often difficult when dealing with multiple plant families (Heckenhauer et al. 2016, H. Korpelainen pers. obs.). Therefore, we decided not to use the universal two-locus plant barcode but chose the ITS region, specifically ITS2.

Previously, Keller et al. (2015), Richardson et al. (2015a) and Sickel et al. (2015) have analysed pollen samples using both microscopy and DNA metabarcoding (the ITS2 barcode) and they found that metabarcoding exhibited higher sensitivity for identifying taxa present in large and diverse pollen samples relative to microscopy. It is also a benefit that metabarcoding does not require a high level of taxonomic expertise. Additional plastid markers, such as matK and trnH-psbA, may provide improved sensitivity to pollen analyses (Richardson et al. 2015b). Despite potential limitations, such as the discrimination capacity of DNA barcodes and quantitative assessment of taxa, and the necessity of specific laboratory facilities and an intensive bioinformatics pipeline, DNA metabarcoding has high potential as an approach to analyse, for example, pollen and fungi present in the environment.

People are exposed to pollen not only outdoors but also indoors. In fact, the question of the remanence of pollen grains indoors is important, because allergic patients could get symptoms from the indoor pollen far away from the pollination period. Besides pollen, indoor air typically contains also other biological particles, such as fungi, Adams et al. (2013) have surveyed temporal variation in fungal assemblages, both indoors and outdoors, using ITS1 pyrosequencing. They discovered that indoors fungal assemblages were strongly determined by dispersal from outdoors. Additionally, there are specific diversity-related considerations, as pointed out by Dannemiller et al. (2014), who demonstrated significant associations between low fungal diversity in indoor air and childhood asthma development in a low-income, Mexican immigrant community in the USA. The indoor air study by Dannemiller et al. (2014) provides support for the so-called biodiversity hypothesis that proposes a connection between biodiversity and allergic diseases that has been provided in several recent investigations (e.g. Hanski et al. 2012; Ruokolainen et al. 2015). An air quality problem may then rather arise from the presence of certain types of pollen than plentiful diversity.

The used method, DNA metabarcoding, is a potentially effective approach to determine the taxonomic composition and diversity of pollen and possible other plant particles, and it may be suitable for pollen monitoring both indoors and outdoors. In this study, great variation in pollen/plant diversities were detected among buildings. Yet, considerable diversities were found both in winter and summer, which shows that substantial amounts of pollen produced in summer enter buildings and stay there throughout the year.

Acknowledgements - We thank the City of Helsinki representatives for providing sampling sites. This work was supported by the Marjatta and Eino Kolli Foundation through a research project 1055 to H. Korpelainen. The experiments comply with the current laws of the country, in which they were performed.

References

Adams, R.I. et al, D. 2013. Dispersal in microbes: fungi in indoor air are dominated by outdoor air and show dispersal limitation at short distances. - *The ISME Journal* **7**: 1262–1273.

Bohmann, K. et al. 2014. Environmental DNA for wildlife biology and biodiversity monitoring. - *Trends Ecol. Evol.* **29**: 358-367.

Brown, E.A. et al. 2015. Divergence thresholds and divergent biodiversity estimates: can metabarcoding reliably describe zooplankton communities. – *Ecol. Evol.* **5**: 2234-2251.

Bruni, I. et al. 2015. A DNA barcoding approach to identify plant species in multiflower honey. - *Food Chem.* **170**: 308-315.

Castelino, M. et al. 2017. Optimisation of methods for bacterial skin microbiome investigation: primer selection and comparison of the 454 versus MiSeq platform. - *BMC Microbiol.* **17**: 23.

CBOL Plant Working Group 2009. A DNA barcode for land plants. – *Proc. Natl. Acad. Sci. USA* **106**: 12794-12797.

318 Chapman, M.D. et al. 2007. Nomenclature and structural biology of allergens. – *J. Allergy Clin.*
319 *Immunol.* **119**: 414–420.
320 Cornman, R.S. et al. 2015. Taxonomic characterization of honey bee (*Apis mellifera*) pollen
321 foraging based on non-overlapping paired-end sequencing of nuclear ribosomal loci. - *PLoS*
322 *ONE* **10**(12): e0145365.
323 Crameri, R. et al. 2014. Fungi: the neglected allergenic sources. - *Allergy* **69**: 176–185.
324 Dannemiller, K.C. et al. 2014. Next-generation DNA sequencing reveals that low fungal
325 diversity in house dust is associated with childhood asthma development. - *Indoor Air* **24**:
326 236–247.
327 Doyle, J.J. and Doyle, J.L. 1987. A rapid DNA isolation procedure for small quantities of fresh
328 leaf tissue. – *Phytochem. Bull.* **19**: 11-15.
329 Galimberti, A. et al. 2014. A DNA barcoding approach to characterize pollen collected by
330 honeybees. - *PLoS ONE* **9**(10): e109363.
331 Hanski, I. et al. 2012. Environmental biodiversity, human microbiota, and allergy are
332 interrelated. - *Proc. Natl. Acad. Sci. USA* **109**: 8334–8339.
333 Hawkins, J. et al. 2015. Using DNA metabarcoding to identify the floral composition of honey: a
334 new tool for Investigating honey bee foraging preferences. - *PLoS ONE* **10**(8): e0134735.
335 Heckenhauer, J. et al. 2016. Universal multiplexable matK primers for DNA barcoding of
336 angiosperms. - *Appl. Plant Sci.* **4**(6): 1500137.
337 Jantunen, J. et al. 2012. Allergy symptoms in relation to alder and birch pollen concentrations in
338 Finland. – *Aerobiol.* **28**: 169-176.
339 Keller, A. et al. 2015. Evaluating multiplexed next-generation sequencing as a method in
340 palynology for mixed pollen samples. - *Plant Biol.* **17**: 558–566.
341 Khansari, E. et al. 2012. Pollen morphology of *Campanula* (Campanulaceae) and allied genera in
342 Iran with special focus on its systematic implication. - *Flora* **207**: 203–211.

343 Korpelainen, H. and Pietiläinen, M. 2017. Diversity of indoor fungi as revealed by DNA
344 barcoding. - *Genome* **60**: 55-64.
345 Korpelainen, H. et al. 2016. Effective detection of indoor fungi by metabarcoding. – *Ann.*
346 *Microbiol.* **66**: 495-498.
347 Kozich, J.J. et al. 2013. Development of a dual-index sequencing strategy and curation pipeline
348 for analyzing amplicon sequence data on the 1MiSeq Illumina sequencing platform. – *Appl.*
349 *Envir. Microbiol.* **79**: 5112-5120.
350 Kraaijeveld, K. et al. 2015. Efficient and sensitive identification and quantification of airborne
351 pollen using next-generation DNA sequencing. – *Mol. Ecol. Res.* **15**: 8–16.
352 Nelson, M.C, et al. 2014. Analysis, optimization and verification of Illumina-generated 16S
353 rRNA gene amplicon surveys. - *PLoS ONE* **9**(4): e94249.
354 Parducci, L. et al. 2013. Molecular- and pollen-based vegetation analysis in lake sediments from
355 central Scandinavia. – *Mol. Ecol.* **22**: 3511–3524.
356 Pichot, C. et al. 2015. Inference of the pollen penetration and remanence into dwellings using
357 seasonal variation of indoor/outdoor pollen counts. – *Aerobiol.* **31**: 315–322.
358 Richardson, R.T. et al.2015a. Application of ITS2 metabarcoding to determine the provenance of
359 pollen collected by honey bees in an agroecosystem. – *Applic. Plant Sci.* **3**(1): 1400066.
360 Richardson, R.T. et al. 2015b. Rank-based characterization of pollen assemblages collected by
361 honey bees using a multi-locus metabarcoding approach. – *Applic. Plant Sci.* **3**(11): 1500043.
362 Ruokolainen, L. et al. 2015. Green areas around homes reduce atopic sensitization in children. –
363 *Allergy* 70: 195–202.
364 Shannon, C.E. 1948. A mathematical theory of communication. - *Bell Syst. Techn. J.* **27**: 379–
365 423, 623–656.
366 Sickel, W. et al. 2015. Increased efficiency in identifying mixed pollen samples by meta-
367 barcoding with a dual-indexing approach. - *BMC Ecology* **15**(20).

368 Taberlet, P. et al. 2012. Towards next-generation biodiversity assessment using DNA barcoding.
369 - *Mol. Ecol.* **21**: 2045-2050.

Figure legends

Figure 1. Percentages (%) of sequences corresponding wild, garden/crop plants, indoor house plants, and non-domestic fruit or other plant material in five buildings, including two university office buildings, two nursery schools and a farmhouse, based on ITS2 sequences. Each pooled sample included sampling of both horizontal and vertical surfaces. Sequencing for January and July 2013 samples were conducted using 454 FLX pyrosequencing, while sequencing for January and July 2014 samples were performed using Illumina MiSeq sequencing.

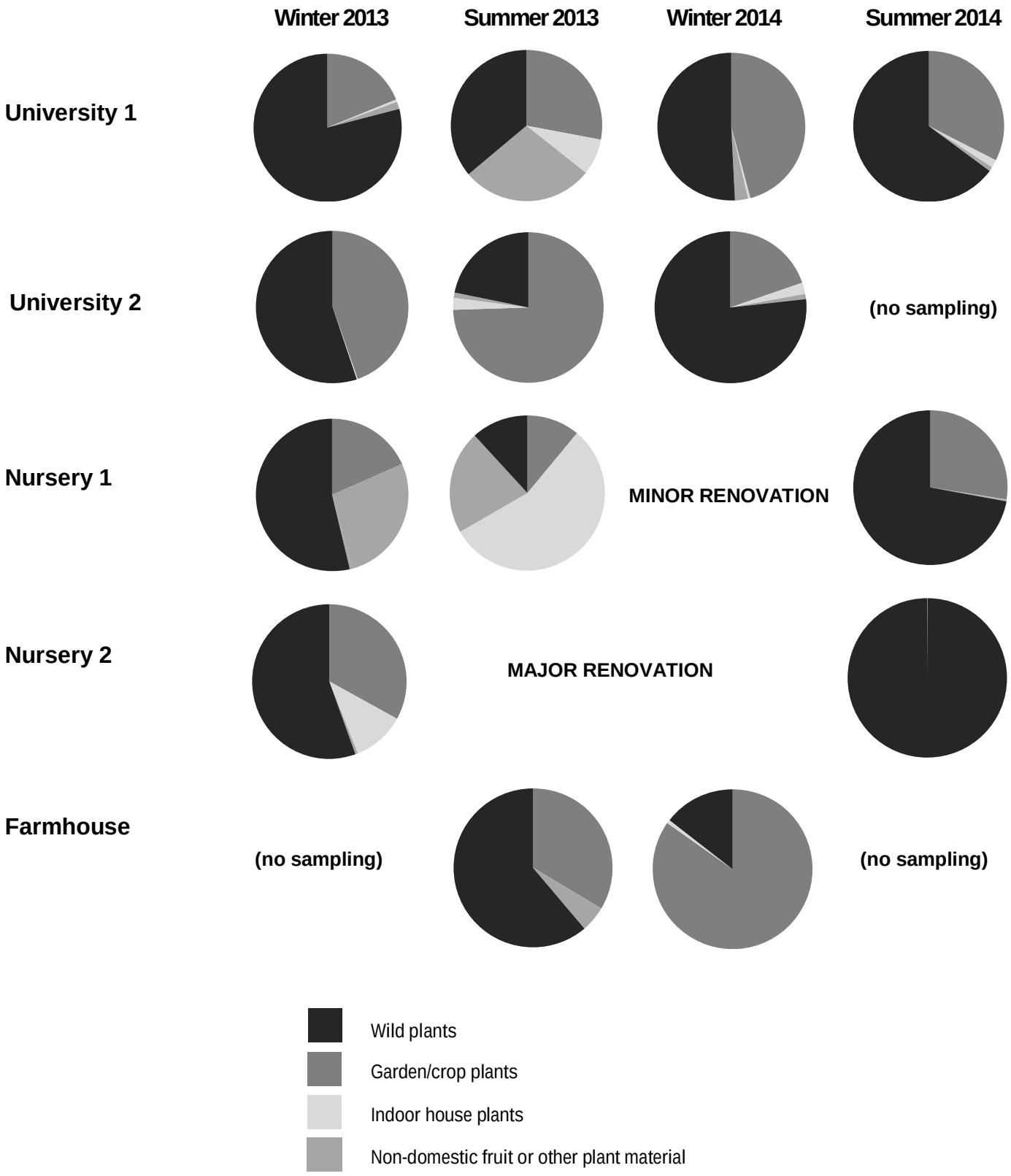
Table 1. Taxonomic diversity of indoor plant material (mainly pollen) at family and genus level in five buildings, including two university office buildings, two nursery schools and a farmhouse, based on ITS2 sequences. Each pooled sample included sampling of both horizontal and vertical surfaces. Range of variables among individual samples is given in parentheses. Sequencing for January and July 2013 samples were conducted using 454 FLX pyrosequencing, while sequencing for January and July 2014 samples were performed using Illumina MiSeq sequencing. N, number of taxa; H, Shannon's diversity index.

		Winter 2013		Summer 2013		Winter 2014		Summer 2014	
Site		N	H	N	H	N	H	N	H
University 1	Family	26 (1-18)	1.63 (0-1.91)	34 (14-27)	2.48 (1.99-2.28)	32 (12-30)	2.50 (1.52-2.52)	40 (22-35)	2.76 (1.86-2.67)
	Genus	45 (1-33)	1.86 (0-1.59)	59 (18-45)	2.95 (2.21-2.69)	44 (12-37)	2.59 (1.38-2.54)	65 (22-58)	3.16 (1.81-3.16)
University 2	Family	9 (4-7)	0.33 (0.05-0.43)	20 (9-16)	1.44 (1.14-2.01)	28 (17-20)	2.33 (1.57-2.08)	(no sampling)	
	Genus	10 (4-7)	0.94 (0.04-0.43)	28 (10-21)	1.48 (0.78-2.38)	49 (22-36)	2.42 (1.59-2.28)		
Nursery 1	Family	18 (4-12)	1.79 (0.83-1.50)	20 (7-11)	1.65 (0.46-1.79)	(minor renovation, no sampling)		28 (11-23)	1.50 (0.14-1.69)
	Genus	25 (4-13)	2.08 (0.83-1.59)	33 (9-16)	1.80 (0.47-2.15)			66 (13-41)	1.59 (0.13-1.91)
Nursery 2	Family	14 (3-10)	2.12 (1.01-1.96)	(major renovation, no sampling)		(major renovation, no sampling)		26 (7-20)	1.03 (0.05-1.82)
	Genus	23 (4-12)	2.53 (1.24-2.09)					40 (7-25)	1.01 (0.04-1.93)
Farmhouse	Family	(no sampling)		25 (4-17)	2.52 (1.35-2.18)	23 (14-19)	0.99 (0.63-2.48)	(no sampling)	
	Genus			39 (4-32)	3.00 (1.35-3.18)	33 (14-25)	0.82 (0.46-2.71)		

Table 2. Five most frequent plant genera and their proportions (%, in parentheses) in five buildings, including two university office buildings, two nursery schools and a farmhouse, based on ITS2 sequences. Each pooled sample included sampling of both horizontal and vertical surfaces. Sequencing for January and July 2013 samples were conducted using 454 FLX pyrosequencing, while sequencing for January and July 2014 samples were performed using Illumina MiSeq sequencing.

Site	Winter 2013	Summer 2013	Winter 2014	Summer 2014
University 1	*Aegopodium* (56.5) *Acer* (11.3) *Mitella* (7.4) *Cannabis* (4.2) *Elymus* (2.8)	*Saintpaulia* (27.7) *Brassica* (9.1) *Dioscorea* (7.0) *Camelina* (4.8) *Juglans* (4.8)	*Betula* (26.1) *Cucurbita* (18.8) *Helianthus* (10.8) *Syringa* (9.3) *Brassica* (6.5)	*Betula* (14.3) *Capsella* (13.3) *Artemisia* (12.4) *Solanum* (11.2) *Brassica* (4.1)
University 2	*Brassica* (55.0) *Coincya* (36.4) *Fagopyrum* (7.9) *Elymus* (1.3) *Hordeum* (1.3)	*Fallopia* (64.2) *Mycelis* (12.9) *Pinus* (4.8) *Arrhenatherum* (3.4) *Saintpaulia* (2.6)	*Myosotis* (31.2) *Betula* (26.7) *Daucus* (7.1) *Brassica* (4.5) *Syringa* (3.6)	(no sampling)
Nursery 1	*Betula* (30.7) *Gerbera* (26.4) *Rubus* (13.2) *Pisum* (8.0) *Lathyrus* 5.7)	*Ficus* (52.4) *Cinnamomum* (18.5) *Pisum* (5.9) *Betula* (4.9) *Saintpaulia* (3.2)	(minor renovation, no sampling)	*Betula* (47.9) *Brassica* (20.0) *Daucus* (18.3) *Pisum* (6.5) *Artemisia* (1.1)
Nursery 2	*Brassica* (28.4) *Betula* (10.1) *Secale* (10.1) *Ficus* (9.2) *Triticum* (6.4)	(major renovation, no sampling)	(major renovation, no sampling)	*Betula* (97.9) *Acer* (0.3) *Brassica* (0.2) *Prunus* (0.2) *Urtica* (0.2)
Farmhouse	(no sampling)	*Saintpaulia* (16.5) *Brassica* (15.7) *Lactuca* (8.4) *Syringa* (6.0) *Polygonum* (5.6)	*Syringa* (83.4) *Betula* (8.5) *Anthriscus* (0.9) *Pinus* (0.9) *Brassica* (0.8)	(no sampling)

Fig. 1.


Winter 2013
Summer 2013
Winter 2014
Summer 2014
University 1
University 2
(no sampling)
Nursery 1
MINOR RENOVATION
Nursery 2
MAJOR RENOVATION
Farmhouse
(no sampling)
(no sampling)
Wild plants
Garden/crop plants
Indoor house plants
Non-domestic fruit or other plant material